DOCUMENT RESUME

ED 265 206 TM 860 067

TITLE	Results of the 1983 Statewide Assessment. Assessment Report 7: An Update on the Alaska Statewide Testing Program.
INSTITUTION	Interwest Applied Research, Inc., Portland, OR.
SPONS AGENCY	Alaska State Dept. of Education, Juneau. Office of Public Information and Publications.
PUB DATE	[84]
NOTE	13p.; For related document, see ED 262 069.
AVAILABLE FROM	Alaska Assessment Materials, Interwest Applied Research, Box 8189-University Station, Portland, OR 97207.
PUB TYPE	Collected Works - Serials (022) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Education; Evaluation Methods; Grade 4; Grade 8; *Mathematics Skills; *Reading Skills; *Standardized Tests; *State Programs; Test Construction; *Testing Programs; Test Interpretation; Test Results
IDENTIFIERS	*Alaska State Department of Education; Assessment Instruments

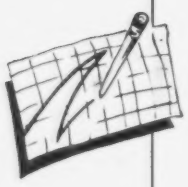
ABSTRACT

The Alaska Statewide Assessment Test is administered to fourth and eighth graders every two years in reading and mathematics. Assessment Report 7 describes how the assessment process works, analyzing results, test content, results for 1983, and statewide average scores. A question-answer format is used to discuss the following: (1) Was students' performance satisfactory? (2) How does 1983 performance compare with 1981? (3) Who scored higher--boys or girls? (4) How does district size affect scores? (5) How well did Alaska Natives do on the test? (6) How might district size relate to Alaska Natives performance? (7) What is the relationship between race and bilingualism? (8) Why are Native students' scores consistently lower? (9) How is performance on one subtest related to performance on another? (10) How good are the Alaska Statewide Assessment Tests? and (11) What impact will these assessment results have on classroom education? Finally, the report discusses interpreting results, plans for 1985, and the 1983 status of Alaskan education. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

JAN 31 1985

ASSESSMENT REPORT 7: An Update on the Alaska Statewide Testing Program



ED265206

Results of the 1983 Statewide Assessment

How well can Alaska students understand what they read? Use simple tables, graphs, or maps to answer questions? Perform one- and two-step mathematics operations? Every other year, the Alaska Department of Education seeks answers to these and other questions through its statewide assessment of fourth and eighth graders.

Alaska's fourth statewide assessment, conducted the week of March 21-25, 1983, was administered to more than 13,000 students—the largest number ever to participate. (Previous tests were administered in 1977, 1979, and 1981.) All tests were scored last spring and reports were returned to schools and districts before the summer recess. Detailed statewide totals were accumulated over the summer.

The Alaska Department of Education publishes results on a statewide basis according to students' sex, race, bilingualism, and district size. Current policy, as set in previous years by the State Board of Education, prohibits comparisons among classrooms, schools, or districts. Although the assessment covers objectives Alaska educators identify as important, some districts may still feel that portions of the test content do not match local goals. Further, the wide variations in enrollment of Alaska districts make accurate statistical comparisons impossible.

What the assessment can do, however, is to give Alaska educators and citizens a chance to compare the performance of local students to that of Alaska students statewide; and then to determine whether those test results should be used in shaping future educational programs. It is hoped that this assessment report will help educators and others begin the important task of objectively analyzing assessment results.

The assessment was administered to over 13,000 students—the largest number ever to participate.

HOW THE ASSESSMENT PROCESS WORKS

Content Panels comprising experts in reading and math from throughout Alaska determined which specific skills would be measured by the 1983 test. Contractors working with the Department of Education then created test items to meet panel members' specifications. Every test item

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

[X] This document has been reproduced as received from the person or organization originating it.

[ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

IM 800 067

was reviewed for suitable content, appropriateness of grade level, and freedom from bias of any kind. Test items were then pilot tested and reviewed a second time prior to printing. In order to permit comparisons between years, the 1983 tests are similar to those of 1981 in format and approach, but almost every item is new.

All tests were administered locally, in most cases by classroom teachers. In early March, participants received tests, machine readable answer sheets, and directions. Tests were divided into two sections, reading and math, each containing about 40 questions and requiring just under an hour to administer. All fourth and eighth graders were tested except those who were absent on the day of administration, those in self-contained special education classrooms, and those in alternative schools. Answer sheets were returned to the Department's contractor for scoring.

The Alaska Statewide Assessment Test is truly Alaska's own. It was constructed by Alaska educators, based on the reading and math objectives they felt were important for Alaska students.

The test offers important advantages over commercially available standardized tests. Not only are the items designed to reflect objectives deemed significant in Alaska, but many contain references to names or places familiar to Alaska students. And because the tests are related to specific objectives, the results can give state and local education leaders more precise information about their students' strengths and weaknesses.

ANALYZING RESULTS

Test scoring was rapid, thanks to a unique scoring system that makes use of a microcomputer and small test scoring machine. Most reports were on their way to local educators the day after the answer sheets were received. Classroom reports, listing individual student scores as well as classroom average scores, were returned directly to principals (or in some cases, district offices) for distribution to classroom teachers. Schools also received information on statewide averages, so that they could compare their local students' performance to that of Alaska fourth and eighth graders as a group.

Since the tests were scored, the Department's scoring contractor, Interwest Applied Research, has been tabulating and analyzing data. This report summarizes their findings.

This report is intended as an objective summary of assessment results. No implications for instructional programming or curriculum development are included or implied. Such decisions properly belong to local Alaska educators, who must consider these findings in light of each district's unique situation, and to the State Board of Education and Department of Education officials who set policy for statewide education.

ABOUT TEST CONTENT

Prior to Alaska's first statewide assessment in 1977, Content Panels of teachers and curriculum specialists began putting together a comprehensive collection of objectives and test items in reading and math from which projects like the statewide assessment could draw. The result of their efforts, the Alaska Objectives and Items Bank (AOIB), forms the basis for the Statewide Assessment Test, the classroom-based Alaska Instructional Diagnostic System (AIDS), and a forthcoming computer-based mastery testing system (now in its design stages) that allows districts to construct reliable tests based on their own objectives.

In selecting items for the assessment test, panelists recognized that each district had its own curriculum and its own subject priorities. Yet they realized it was important to know how students across the state performed in reading and math. They decided to focus on basic skills, but not just at the minimum level; the assessment test contains both items which they felt every student should master and items covering more advanced skills which, though important, might not be part of every curriculum. Because of this mix of basic and advanced skills, the statewide assessment should not be viewed as a "minimum competency" test (a test which measures only very basic skills, usually at several grades below the students' level).

The test offers several advantages over commercially available standardized tests; because the tests are related to specific objectives, the results can give precise information about students' strengths and weaknesses.

The following sections describe each of the six assessment subtests, and present sample items for each.

MATHEMATICS

Application

Application items test problem solving skills: identifying exact facts, translating written information into math equations, determining the correct operation to solve a problem, recognizing when more information is needed. Most items consisted of a short story problem, with one or two relevant questions.

Example: (Grade 8)
A tour group was visiting the Mt. Hood Glacier. There were thirteen men, fifteen women and eight children. How many were in the tour group?

What is the way to find the answer to this problem?

a. Add
b. Add, then subtract
c. Subtract
d. Subtract, then add

Computation

Computation involves basic arithmetic: adding, subtracting, multiplying and dividing whole numbers (positive and negative), fractions, and decimals. The problems were not intended to be tricky, yet were constructed in such a way that common errors would lead students to select one of the incorrect options. Such a test format helps educators who interpret assessment results to make an educated guess about the cause of student errors.

Example: (Grade 8)

$$\begin{array}{r} 32.2 \\ \times\ 1.5 \end{array}$$

a. 4.72
b. 4.83
c. 47.2
d. 48.3

Concepts

Test content and format have changed since the assessment began. Between 1979 and 1981, the Content Panel determined it was important to test students' understanding of math concepts—such things as number properties, place value, rounding, telling time, units of measurement, equivalent fractions, geometric shapes. Math concept items measure a student's understanding of information presented, but do not require any calculation.

Example: (Grade 4)

The time 7:05 means

a. Seven minutes before five o'clock
b. Five minutes after seven o'clock
c. Five o'clock
d. Half past seven o'clock

READING

Comprehension

For this section of the test, students were given a reading passage, then asked to respond to several questions about that passage. The questions were designed to measure such skills as ability to identify the main idea, distinguish facts from opinions, draw inferences, and reach conclusions. Great care was taken to ensure that the reading level of each passage was appropriate for the grade level tested.

Example: (Grade 8)

The following information was provided by the American Automobile Association.

The American Automobile Association (AAA) has recently produced a new map of Alaska and Northwest Canada. This map, which is the first complete revision made by AAA in twenty-five years, required 4,000 hours of work to develop. It is based on data obtained from United States survey maps, from satellite photographs and information collected from first hand observation of mileage and travel conditions.

Printing and designing the map also required many hours of work. It was printed by a four-color process and has eight smaller strip maps.

A spokesman for AAA's travel department informed us that this is the only accurate road and travel map of the area one can obtain.

From the information given, what would be the BEST guess about the amount of time that went into the project?

a. One day
b. Two or three weeks
c. More than a year
d. Ten years

Study Skills

This section focuses on students' ability to find and selectively use information in a dictionary, index, map, graph, chart, table, and so on. Special care is needed in interpreting results from this section, since some study skills covered on the assessment test may not be part of some Alaska districts' curriculum.

EXAMPLE: (Grade 4)

(Student is given a sample dictionary page on which to base responses.)

Which of these statements is NOT true?

a. A book would be easy to MOLD.
b. MOLES live underground.
c. Volcanoes give off enough heat to form MOLTEN rock.
d. MOLASSES tastes very sweet.

Word Skills

This section of the test covers skills Content Panelists consider vital to reading proficiency, including knowledge of regular and irregular vowel sounds, consonant combinations, possessives, prefixes and suffixes, synonyms and antonyms, and multiple meaning words.

EXAMPLE: (Grade 4)

Read this sentence and answer the question.

The doctor CHECKED her weight.

What does CHECKED mean in the sentence?

a. Covered with black and white squares
b. Marked with a pen or pencil
c. Measured
d. Ordered to pay money through a bank

HOW HARD WAS THE TEST?

Every test has an easiest and a hardest item—a "snap" item that practically no one misses and a "toughie" that only a few get right. The 1983 "Toughie Trophy" goes to item 30 of the grade 8 math test. This item, which asks students to recognize that if $3/A = B$ then $A = 3/B$, was answered correctly by only 28 percent of the students. The easiest item in 1983 was also on the grade 8 math test. When asked how to find the answer to the word problem (given previously as the example math application problem) which involved totalling the number of people in a tour group, over 97 percent of the students knew it was necessary to add the numbers to solve the problem.

Because the Content Panels were interested in learning the performance of students on a wide range of reading and mathematics skills, such variations are perfectly acceptable. Unlike some standardized tests which must be constructed with a limited range of difficulties, the objectives-based Alaska Assessment Tests are able to measure student performance on any skill which is considered important by the state's educators.

RESULTS FOR 1983

The graphs in the following section illustrate the statewide averages for fourth and eighth graders on each of the six subtests. Notice that results for the two grades are graphed separately. There is an important reason for this. The test items for the two grade levels are not equal in difficulty. Even though fourth graders scored "higher" than eighth graders on the Math Concepts subtest, for instance, it is not correct to say that they performed "better," since the two subtests were not comparable. The items that the Content Panel members chose for the grade 8 test may simply have been more difficult.

Remember that the underlying purpose of assessment is to identify instructional needs for the purpose of improving programs and teaching strategies. For this reason, the scores are most valuable when they are compared against expectations held for the students by Alaska educators and the general public, rather than when viewed in isolation.

STATEWIDE AVERAGE SCORES

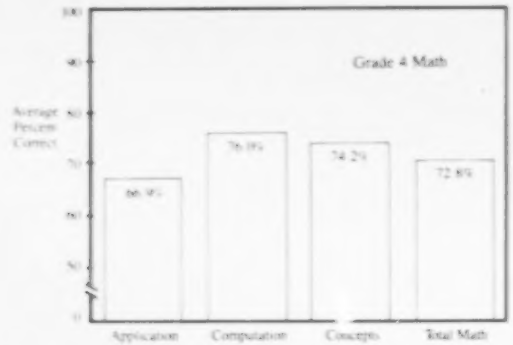
These graphs show the statewide average scores obtained by Alaska fourth and eighth graders on each of the six subtests. Based on this review of performance, some initial observations can be made.

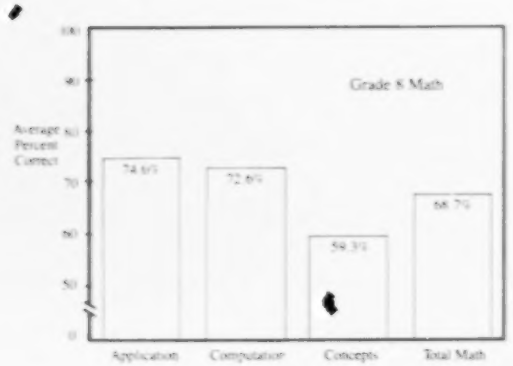
Mathematics

The scores on the math subtests were clustered around the 75 percent correct level with two important exceptions. The grade 4 math application score was a little less than 67 percent; this score is somewhat lower than the equivalent 1981 score due primarily to lower performance on one item related to obtaining a fact from the given word problem. The grade 8 math concepts score (59%) is virtually identical to the 1981 score. This subtest requires the students to show an understanding of some advanced mathematics concepts such as pi, number properties, geometry and so on. Students scored rather poorly on some of these items (as low as 28%) and this lowered their overall subtest score. It is likely that the difficult items are not emphasized in all schools by the time the assessment test is given.

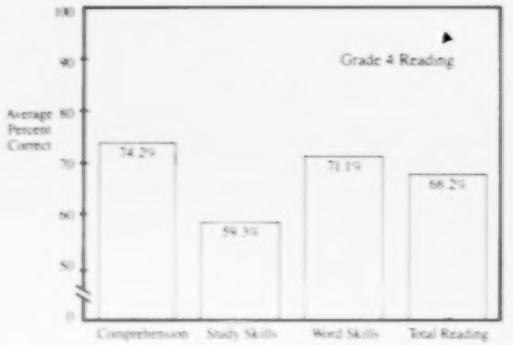
Eighth grade students raised their math computation scores by over four points from 1981, but part of the im-

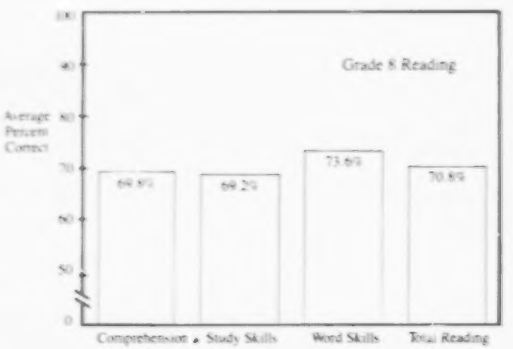
Math





Reading





provement may be due to a change in a few items. Two items dealing with negative numbers were thought by the Content Panel members to have little practical significance and were therefore replaced. The newer items, while still not easy, were answered correctly by substantially more students.

Reading

Scores on the reading section of the test varied considerably depending on grade level and subtest. Overall, the total reading scores were quite similar (about 70% correct) but the contributions made by the three subtests differed considerably. Fourth graders, for example, needed high scores in reading comprehension (74%) and word skills (71%) to make up for a low score in study skills, where they answered only 59 percent of the items correctly. As with the eighth grade math concepts test, this lower performance may be attributable to the fact that some items reflect objectives that are not completely covered by all districts prior to test administration.

The grade 8 reading scores were quite consistent. While the highest score (74%) occurred in word skills, this is quite close to their lowest score, the 69 percent in study skills. Remember that in a 10-to-15 item subtest, answering one more item correctly will raise a score from eight to ten percentage points. So the five percent difference between highest and lowest subtest (74% and 69%) means that the total number of items answered correctly was less than half an item different between word skills and study skills.

If Alaska educators were asked to predict the subtests in which students would score lowest, most would probably correctly pick math concepts and reading study skills. For one thing, there is no established curriculum for teaching study skills or math concepts. Although the Content Panels believe these skills to be very important, they recognize the disparity among Alaska districts in the manner and extent to which this material is covered.

REMEMBER THE HOTLINE!

Do you have a question about the assessment that isn't covered here? A suggestion? Criticism? Recommendation for improvement? The Department of Education is eager to encourage your participation, and for the past several years has made a HOTLINE number open for just this purpose. Please call ***COLLECT*** any weekday from 9 a.m. to 4 p.m. Pacific time, and someone knowledgeable about the assessment will be happy to talk with you.

ASSESSMENT HOTLINE
503-223-3396 COLLECT

LET US HEAR FROM YOU!!

ANSWERS TO A FEW IMPORTANT QUESTIONS

An assessment can only be useful when educators and others understand what the results mean. Certainly not every question about Alaska's 1983 assessment can be answered here, but an effort has been made to respond to those questions most often considered critical. Please remember that far more detailed information can be obtained either through the Alaska Department of Education in Juneau, or its contractor, Interwest Applied Research in Portland, Oregon.

Was students' performance satisfactory?

This is perhaps the single most difficult question to answer in regard to any assessment. "Satisfactory" performance is really a subjective judgment, which must take into account (1) the difficulty of the test items, and (2) the expectations held for the students who took the test.

To understand this better, suppose that in a given School District X, every item on the test (in other words, all test content) had been thoroughly covered in classroom instruction. Further, suppose that the staff in that district felt their students were capable of mastering all those skills and had, indeed, done so prior to the test. They might then reasonably expect rather high performance from District X students.

But suppose that in District Y, some of the test content had not been covered during classroom instruction, and, as a result, the staff in the district had very modest expectations for their students. One would not, therefore, expect the performance of District Y students to be as high as those in District X. If their scores were identical, staff in District X could feel the performance was unsatisfactory, while the District Y staff might feel the same scores were quite satisfactory.

The staff and parents in individual districts are best qualified to decide what level of performance is "satisfactory," since obviously they are best acquainted with local objectives and the expectations held for local students. At the same time, some general observations can be offered on a statewide basis. A special Interpretive Panel appointed for this purpose will soon initiate an item-by-item review of test performance (see the discussion later in this newsletter for details). In making their judgments, the Interpretive Panel will consider which test items represent minimum skills, or

Satisfactory performance is really a subjective judgment which must take into account (1) the difficulty of the test items, and (2) the expectations held for the students who took the test.

those which most Alaska students at each respective grade level should have mastered; and which represent more advanced skills, or those which fewer students would have mastered. In addition, they will rate the relative difficulty of each individual item and their expectations for overall student performance before offering any interpretations of results.

How does 1983 performance compare to that for 1981?

This frequently asked question about year-to-year changes in test performance cannot always be answered through a simple comparison of test scores. Such data could be very misleading, since tests for any two years could be somewhat different. Moreover, other factors are likely to have changed to some degree: areas of statewide educational emphasis, classroom and district objectives, and so on. Direct score-to-score comparisons, therefore, must be made cautiously.

Such a comparison is possible between the 1981 and 1983 statewide assessment tests, however, because the 1983 Content Panels thought that the objectives used in 1981 were still important indicators of statewide student learning. Under their direction, the 1983 test was made very similar in content to that administered in 1981. Further, no major changes in educational emphases occurred at the statewide level between 1981 and 1983. Thus, comparisons of student performance for 1981 and 1983 can be made with some degree of confidence.

At the same time, despite comparable content, the 1981 and 1983 items are not exactly equivalent in difficulty. Even items measuring the same skill can have slight differences. For this reason, comparisons made at this time must not be too precise; a change of two or three percentage points, for instance, should not be regarded as significant. Each subtest has between 10 and 16 items; an increase of three percent is an improvement of less than half an item answered correctly.

Given the precision of the tests, one can say that 1983 scores are little changed from 1981. Some subtest scores went up, some went down. But the overall conclusion based on comparison of the 1981 and 1983 test scores is that the performance of Alaska students is little changed.

These similarities are not surprising. In the two intervening years, there were no significant statewide educational changes that would have led to higher assessment scores. In general, statewide scores will improve

The conclusion is that the performance of Alaska students is little changed from 1981. Statewide scores will improve dramatically only if increased emphasis is given, on a statewide basis, to the skills covered by the test.

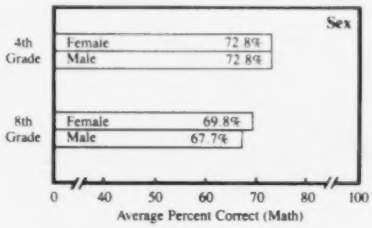
dramatically only if increased emphasis is given, on a statewide basis, to improving mastery of the skills covered by the test. This does not mean, however, that scores did not change in individual districts. It is likely that some districts did make instructional changes that were reflected in increased assessment scores.

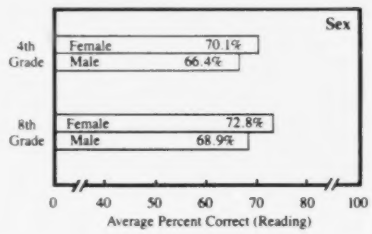
Who scored higher—boys or girls?

As in the past Alaska assessments, girls slightly outscored boys on most test sections. Fourth grade girls averaged 31.5 items correct on the 45-item reading test and 27 items correct on the 37-item math test. Boys averaged 30 items and 27 items correct on reading and math respectively.

At the eighth grade level, girls averaged 30.6 items correct on the 42-item reading test, and 30 items correct on the 43-item math test. Eighth grade boys averaged 29 items on both the reading and math tests.

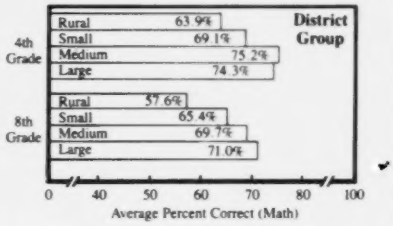
Although girls slightly outperformed boys, the differences probably have little educational significance. In terms of number of items answered correctly, the differences between boys and girls are slight. There are probably hundreds of reasons for these differences, and they cannot be determined through the information collected during this assessment. There is a high probablity that the most important factors are social and only partly related to instruction or classroom practices.

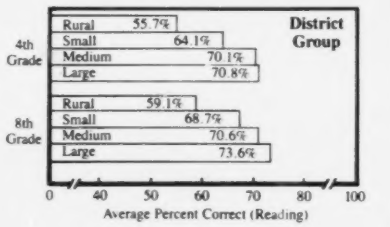




How does district size affect scores?

For purposes of this comparative analysis, Alaska's 53 districts were categorized by size and type—the 2 **large** districts being those with over 10,000 students and the 6 **medium** districts those with 1500 to 10,000 students. The 24 other city districts were classified as **small**, and the remaining 21 (all outside incorporated areas) were classified as **rural**.





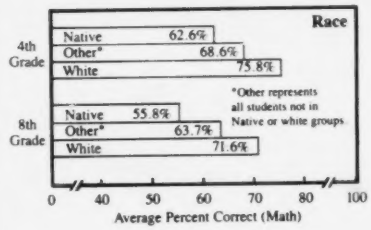
At the eighth grade level, rural districts received the lowest scores and, as the accompanying graph shows, scores improved as district size increased. At the fourth grade level, the same pattern emerges with the exception that students in medium sized districts slightly outperformed their peers in large districts on the math test.

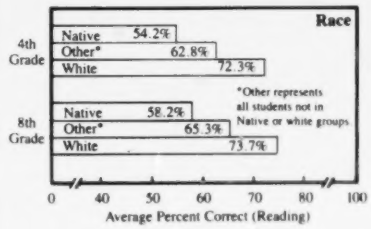
District size alone cannot be said to account for differences in performance, however, since several small and rural districts had scores far above the state average. Moreover, district size is often related closely to other factors such as isolation of the schools and experience of the teaching staff. There are also differences in the characteristics of the students attending small and rural districts. In short, the size of districts is related to performance on the assessment test, but small size is probably not, by itself, a reason for lower scores.

How well did Alaska Natives do on the test?

Alaska's Native students scored substantially lower than non-Natives on all subtests. Fourth graders averaged 7.7 fewer items correct on the reading test and 4.6 fewer items correct on the math test. Eighth graders averaged 6.2 fewer items correct on the reading test and 6.6 fewer items correct on the math test.

The educational significance of these differences is a matter for the Interpretive Panels, the State Board of Education and district level educators to explore further. Meanwhile, those considering the results of this assessment should keep in mind the potential for interaction between race and other factors: geographic location, for instance; access to resources; experience in taking tests; bilingualism; and other factors.





The OTHER category shown on the graph includes blacks, Hispanics, Asians, and other non-white students. Given the highly diverse characteristics of this group, few if any inferences should be drawn regarding their performance.

How might district size relate to Alaska Natives' performance?

Native students form a higher proportion of the student population in small and rural districts than in medium and large districts. It has been suggested that attendance at smaller schools, where resources may be limited, could account for some of the difference in performance between Native and non-Native students. How-

ever, statistical analyses based on this premise fail to account for the total difference. White students outscored Natives regardless of whether the Native students were in urban or rural schools. Nevertheless, Natives in urban schools scored higher than those in rural schools.

Additional, more thorough analyses could be undertaken to determine the relationship between race and district size. Such variables as bilingualism of teachers and students, length of enrollment in a particular type of school or district, expectations of parents and teachers, experience of teaching staff, attendance patterns, and differences in district or school objectives and teaching approaches would have to be considered. However, the complexity of the interactions between factors make the relationship between district size and Native performance difficult to determine precisely.

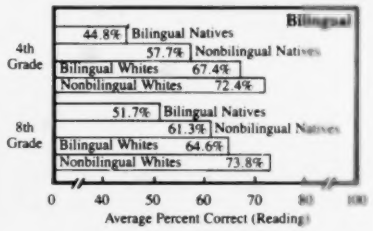
What is the relationship between race and bilingualism?

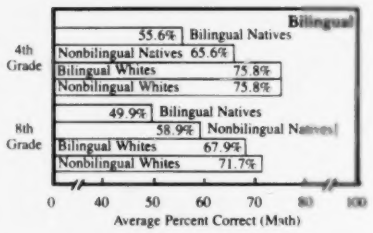
Bilingual Native students scored consistently lower than Native students who spoke only English. For example, fourth grade bilingual Natives averaged 45 percent of items correct on the reading test, 56 percent correct on the math test. Nonbilingual Natives (those speaking English only) averaged 58 percent of items correct on the reading test, 66 percent correct on the math test.

At the eighth grade level, bilingual Natives averaged 52 percent of items correct on the reading test, 50 percent correct on the math test. Natives who spoke English only averaged 61 pecent of items correct on the reading test, 59 percent correct on the math test.

Bilingual white students at both fourth and eighth grade levels scored lower than nonbilingual whites except in fourth grade math, but higher than either bilingual or nonbilingual Natives. Since there were only 59 fourth grade and 35 eighth grade white bilingual students tested, however, these results are not as reliable as the others reported in this newsletter and so should not be overemphasized.

Like other factors influencing performance, bilingualism is affected by a number of variables. In order to properly analyze these findings, it is important to ask questions like the following: Does the bilingual student speak English as a first or second language? What is the degree of fluency in the first and second language, and are they approximately equal? To what extent, if any, has bilingual instruction been provided? And suppose the tests had been administered in a language other than English; what differences in performance might have occurred?





Why are Native students' scores consistently lower?

Countless factors can affect test performance, but not all of them can be learned through research. It is extremely important for those reviewing the results of the Alaska Statewide Assessment to realize that being Native does not in and of itself ***cause*** a lower score, but it may relate—in both obvious and subtle ways—to numerous factors that can affect test performance. Two important variables—bilingualism and district size—have already been briefly considered. What other factors might be significant?

First, many of Alaska's Native students reside in remote areas amid harsh environments. Some may not be familiar with testing formats and purposes of testing. Further, some Native students may not equate success on a written test with "achievement" in the same way that a student from another environment has been conditioned to do. Second, access to educational resources may play a role. Students in rural areas typically have available fewer magazines, books, newspapers, and so on. Communications through television and radio are more limited. Differences in lifestyle can account, to a very great degree, for the opportunities one has to develop numerous everyday math, vocabulary building, and reading skills.

Third, in cases where Native students are bilingual but instructors are not, instruction may be hindered to a degree because of problems with instructional or interpersonal communications. This situation is further complicated by the fact that instructor turnover in rural areas is typically high. However, only further research—research beyond the scope of the assessment—can really determine whether these, or any of a dozen other factors, have significant impact on assessment scores.

The magnitude of the differences between Native and non-Native students makes it clear that—whatever the reasons—the performance of Native students on the assessment tests is substantially lower.

Readers should also recognize that the whole relationship of cause and effect is a particularly tricky one to analyze with respect to test scores. Hundreds of factors may be related in some way to test performance, but cannot really be said to ***cause*** high or low scores. For example, if it happened on a particular math test that all left-handed students outscored right-handed students, statistics might show an apparent relationship between left-handedness and test performance. But

clearly, it would be misleading to conclude that left-handedness ***caused*** superior performance—even though the evidence might seem to indicate that. Statistical significance does not always mean educational significance. The point is, statistical test data must be carefully analyzed and interpreted in light of other factors, such as knowledge of students and their learning environment.

Much more evidence is needed before the reasons for differences in Native and non-Native performance can be outlined with any precision. But the magnitude of the differences between Native and non-Native students makes it clear that—whatever the reasons—the performance of Native students on the assessment tests is substantially lower than that of non-Native students. The responsibility now falls to Alaska educators to determine what steps, if any, should be taken to change the imbalance.

How is performance on one subtest related to performance on another?

Students who did well on one reading subtest tended to do well on the other two. This pattern held true for math as well. At the fourth grade level, the strongest relationship was between Reading Word Skills and Reading Study Skills; students who scored high on Word Skills were likely to score high on Study Skills as well. At the eighth grade level, equally strong relationships were found between the Math Computation and Math Concepts subtests and the Reading Comprehension and Reading Study Skills subtests.

It has often been suggested that math story problems may test reading comprehension as much as or even more than math skill. If that were true, one would expect a strong correlation between math application and reading comprehension subscores. The data analysis did show a high correlation between these two subtests at the eighth grade level, but at the fourth grade the relationship was only modest. As a general rule, students who did well on one subtest were, not surprisingly, likely to do well on other subtests, regardless of the content.

How good are the Alaska Assessment Tests?

Since the Assessment Program began eight years ago, hundreds of Alaska educators have helped in establishing objectives, determining test content, setting assessment policies and reviewing test items. These educators can take considerable pride in the reading and math assessment tests they have worked to develop. Since the earliest version was piloted in 1976, all Alaska assessment tests have been carefully evaluated for two factors which best measure a test's quality: reliability and validity.

The Alaska Assessment Tests have demonstrated validity for statewide testing purposes. And analyses indicate that their reliability is excellent.

Reliability is a measure of consistency. If two groups of students with comparable skills scored quite differently on a given test, that test would be said to have low reliability. Similarly, if a given group of students scored very low on a test at one administration, then—without any additional instruction—scored very high at a second administration, that test would also be said to have low reliability.

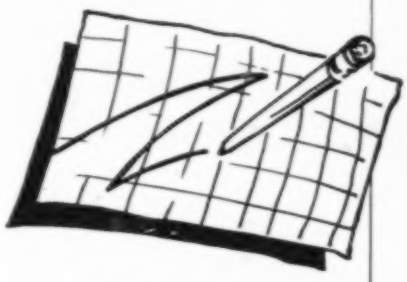
Validity, which is most often determined through review of the test items by knowledgeable educators, indicates whether a test in fact measures what it purports to measure. For example, if items on a psychology exam required very high level reading skills, the test might well be more a measure of reading comprehension than of psychological theory—and thus would be said to have low validity.

The Alaska Assessment Tests have demonstrated validity for statewide testing purposes. And analyses indicate that their reliability is excellent—fully equivalent to that of nationally published achievement tests used by many districts. And of course, because the Alaska Assessment Tests are developed and reviewed by teams of Alaska educators, they offer the additional advantage of content reflective of students' experience and interest.

What impact will these assessment results have on classroom education?

The answer to this question will likely vary from district to district—even from classroom to classroom. Local educators must ask to what extent the tests relate to local objectives; in cases where the relationship is strong, test results can be used to determine how well educational goals are being met, and whether reading or math programs need revision or expansion.

Some districts stress objectives not covered by the assessment tests. The quality of instruction on skills not covered by the assessment obviously cannot be determined by the test results. There may also be a few instances where the assessment tests cover skills not included in the districts' objectives. It is unlikely that any district would find large amounts of test content that is unrelated to their own objectives. Large discrepancies might lead districts to consider whether test results suggest that any change in educational emphasis is warranted.



Many Alaska educators agree with the view of the Conent Panels that the math and reading skills tested in the assessment provide an important foundation upon which Alaska students' academic success depends. So in general, educators in local districts will find the scores from the assessment useful—particularly when compared to local expectations.

An important role of the Statewide Assessment is to provide information which is not otherwise available about the performance of Alaska students on some specific objectives important to their education. Groups such as the State Board, the Department of Education, and teachers' and administrators' professional organizations may use the statewide results to determine what changes, if any, might be made to improve education in Alaska.

Where can I get more information?

Far more detailed information is available than can be presented in this newsletter. If you would like to know more, please contact Al Hazelton at the Alaska Department of Education, Pouch F, Juneau, Alaska 99811, 465-2966. Or contact Evelyn Brzezinski at Interwest Applied Research, Box 8189—University Station, Portland, Oregon 97207, (503)223-3396. Either person will be pleased to answer your questions about the 1983 assessment, send you copies of the tests, or provide information about scoring, reporting, or test review procedures.

An important role of the Statewide Assessment is to provide information which is not otherwise available ... Groups may use the statewide results to determine what changes, if any, might be made to improve education in Alaska.

WHAT IT ALL MEANS: INTERPRETING THE RESULTS

The next step in the assessment process is determining how students' performance relates to the expectations of educators across the state. This task falls to the approximately 70 teachers, curriculum specialists, and educational administrators from throughout Alaska who have been selected to serve as this year's Interpretive Panel members. These 70 persons were nominated last spring by Alaska district superintendents and by the District Testing Representatives who assisted with the test administration and reporting in 1983.

The next step ... is determining how students' performance relates to the expectations of educators across the state.

Interpretive Panel members—who will begin working within the next month—will use their knowledge and expertise to set two performance levels for each test item. The first level will indicate the "minimum" percentage of students that panel members feel must get an item correct in order for performance to be considered acceptable. For example, a minimum level of 60 would suggest that 60 percent, or more, of all students tested must get a given item right in order for performance on that item to be regarded as acceptable.

Panelists will also set a "desirable" level for each item. A level of, say, 80 would indicate that panelists would like to see 80 percent, or more, of all students tested answer a given item correctly. The desirable level represents the goal Alaska educators hold for instruction related to the skill covered by the item.

In setting these performance levels, panelists must consider the relative difficulty of each item and whether an item measures basic or advanced skills.

First the Expectations—Then the Results ...

Panel members have no advance knowledge of actual student performance when they set their expectation levels. After they have completed their item-by-item review, staff at Interwest Applied Research in Portland will compare actual results to the panelists' performance levels for each item.

Wherever performance falls below the "minimum acceptable" level, it will be said to indicate a weakness in student performance. Performance above the "desired" level will be said to indicate a strength. The performances above the "minimum" but below the "desired" level will be regarded as satisfactory.

These results will be included in the next **Assessment Report** newsletter, to be released by the Department of Education in December.

District Level Interpretation

District educators must keep in mind that the strengths and weaknesses that result from the Interpretive Panel's review are based strictly on statewide performance. The **Assessment Report** newsletter will not address strengths and weaknesses at the district level; thus it would be a mistake to base any local curriculum changes on these interpretations.

Districts can, however, follow the same process for setting minimum and desirable levels of performance, based on knowledge of local objectives and students' educational environment. Free copies of tests are available from the Department for this purpose. Information related to districtwide performance on each test item and an Apple II computer program for conducting a local Interpretive Panel review are also available from the Department of Education. Once strengths and weaknesses in performance have been identified relevant to local expectations, meaningful educational planning can occur.

PLANS FOR 1985

The planning, development, and revision that go into creating the Alaska Assessment Tests are a continuous process — and work on the 1985 tests has already begun. The Content Panel members will follow the same process of selection of objectives and review of items as used in previous years. It is expected that enough continuity in format and content will be retained from 1981 and 1983 so that comparisons can be made among 1981, 1983 and 1985 test performance. At the same time, test developers and Content Panel members will consider whether any improvements can be made in test content, administration and scoring procedures, and so on.

When the test content and format have been established, a new set of items will be developed. As always, each item will be scrutinized to ensure validity, reliability, and freedom from bias. Following this careful review, all items will be pilot tested, and only those of the highest quality retained.

In 1983, the scoring and reporting procedures for the test were improved so that the information would be more useful to local educators. Some additional changes may be made prior to the 1985 assessment to continue this trend. It is possible, for instance, to provide local educators with information on how each student, classroom, and school compared to others in the state. It may also be practical to make comparisons with national averages. Some people have requested these types of information; the Department is now discussing the merit of the comparisons with educators across the state. If you have a preference regarding expansion of the information available, please call the Department of Education to express your opinion.

ALASKA EDUCATION: ITS STATUS IN 1983

A Report to Alaska Educators and Citizens

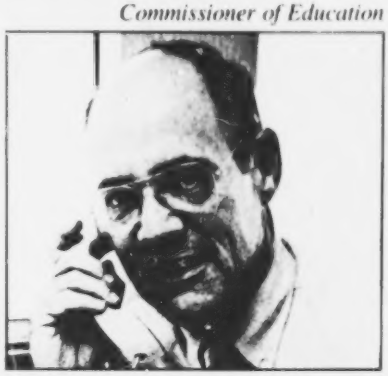
Accountability has become a much discussed topic in public school education. Alaska citizens can be proud of the fact that, over ten years ago, their educators began working on a process for determining how effectively the state's schools were educating our young people. As a result of those efforts, the Alaska Statewide Assessment Program was established. Information obtained from regular assessments of student skills has led to improved learning opportunities for the state's students.

Last March, over 13,000 of Alaska's fourth and eighth grade students participated in the fourth biennial assessment. The tests covered objectives important to the ongoing education of Alaska students, as established by panels of educators from throughout the state. Scores from the Assessment tests are the only information we have from all schools in the state to let us know how well students are mastering basic skills.

Over the past few months, extensive analyses have been performed to show how well students performed on the tests. This Assessment Report presents those results, and highlights plans for further interpretation of the test results. The information presented in this and forthcoming newsletters will be of use to the state's educators as they consider the results of the 1983 testing.

Although I came to Alaska after administration of the 1983 Assessment, I have learned that this important testing could not have been conducted without the assistance of several hundred educators from throughout the state. On behalf of the Department of Education, I wish to thank each of you who participated. We at the Department are eager to make the Alaska Statewide Assessment a valuable tool for understanding and improving Alaska schools. For example, discussions are currently underway to determine if changes should be made in the reporting process, perhaps to include more comparative information among Alaska districts or comparisons with national performance on similar content. The test content for the 1985 testing also is under review. As we consider these and other issues, I would like to solicit your continued interest and support in this process. Any time you have questions or suggestions about the Assessment, please contact me or project director Al Hazelton.

Harold Raynolds
Commissioner of Education



Office of Research & Evaluation
Alaska Department of Education
Pouch F
Juneau, AK 99811

PRESORTED
FIRST CLASS
MAIL
U.S. POSTAGE
PAID
JUNEAU, AK
PERMIT NO. 224

ASSESSMENT REPORT — Please Circulate To All Interested Individuals